Ekoist: Journal of Econometrics and Statistics, 37, 199-211

DOI: 10.26650/ekoist.2022.37.1145052
http://ekoist.istanbul.edu.tr

İSTANBUL ÜNİVERSİTESİ YAYINEVİ

Ekoist: Journal of Econometrics and Statistics

Received: 18.07.2022
Revision Requested: 20.07.2022
Last Revision Received: 20.07.2022
Accepted: 09.08.2022
Online yayın: 14.10.2022

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Association Rules Mining on Retail Data

Hatice Dağaslanı*, Özlem Deniz Başar**

Abstract

The development in information technologies, artificial intelligence, and data mining benefits people in many areas. With this development, data stacks are formed through the storage of ever-increasing data. Accessing useful information from the data heaps is a very difficult process. This has led to the emergence and development of the concept of data mining. In this study, the relationship between the categories of the products sold by a company in the retail sector operating in Turkey was analyzed using the Apriori algorithm, which is an algorithm used in data mining. In the application, one-day sales data of the company was used. The data obtained was provided to extract the association rules with the help of Python. In this way, the purchasing habits of customers were determined by finding meaningful relationships between products using association rules.

Keywords

Apriori Algorithm, Association Rules Analysis, Data Mining

* **Corresponding author:** Hatice Dağaslanı (Master Student), İstanbul Commerce University, Natural and Applied Sciences, Department of Statistics, Istanbul, Turkiye. E-mail: hatice.dagaslani@istanbulticaret.edu.tr ORCID: 0000-0002-2190-4135

** Özlem Deniz Başar (Prof. Dr.), İstanbul Commerce University, Graduate School of Natural and Applied Sciences, Department of Statistics, İstanbul, Türkiye. E-mail: odeniz@ticaret.edu.tr ORCID: 0000-0002-9430-8975

To cite this article: Dagaslani, H., & Basar, O. D. (2022). Association rules mining on retail data. *EKOIST Journal of Econometrics and Statistics,* 37, 199-211. https://doi.org/10.26650/ekoist.2022.37.1145052

Introduction

Data science enables businesses to make good decisions and generate insights that can be used to create innovative structures in products and services. Statistical methods and various algorithms are being developed to transform large amounts of data from meaningless to meaningful. With the application of these algorithms, an attempt is made to obtain meaningful results. Machine learning models are used to facilitate decisions for businesses by obtaining meaningful results.

Machine learning models on the detailed behavior of consumers contribute to the analysis of insights. Thus, strategies can be developed to reveal a stronger understanding of purchasing decisions of customers. In this context, one of the most used methods to analyze customer behaviors is the association analysis method. With the association analysis method, risk analyses can also be done on the habits of customers, and especially for risk management in the banking sector, association analysis rules are applied (Kalikov, 2006: 10).

Firms have large amounts of data of varying nature about their customers. The information obtained by the companies is important in terms of their competitiveness. The techniques used in data mining are from data sets; they are used for obtaining useful information easily. Methods such as clustering, classification, association rules and estimation, aim to extract data that companies can apply under the scope of data mining.

By examining the shopping carts that reflect the shopping habits of the customers, and analyzing which products are bought together, it is possible to update and arrange the aisles and shelves accordingly. With this technique, the products purchased by the customers are analyzed. Thus, products that will increase sales are offered as advice to customers, with suggestions that will be beneficial according to their shopping carts. In this way, the purchasing habits of customers can be determined by finding meaningful relationships between the products sold using association rules. Examining the movements of customers can provide managers with effective marketing development opportunities (Aksoy, 2019: 64).

The analysis of the products purchased by customers can be used as a motion point in cases where it is generally composed of commercially meaningful data, but it is not known which relations to look for on the data set. In this way, various campaigns can be used to increase sales opportunities (Karagöz, 2007: 26).

In the application part of the article, one-day sales data of a retail company in Turkey was used in a five-day period. Due to confidentiality principles, the name of the company will not be disclosed. Purchasing movements of customers shopping for products in more than one category out of 2,521 invoice transactions were examined. The association analysis of the purchased products on the dataset with 1,528 customer (single) transactions was carried out with the help of Python.

Literature

Timor et al. (2011) applied clustering and association rules analysis by examining customer shopping movements in practice in the retail sector. The application was carried out using the SPSS program. According to the result analysis, they suggested that the production should be arranged in accordance with the sales trends, taking into account the company-specific market focus strategy, marketing and awareness activities.

In their study, Söylemez et al. (2016) examined the traffic accidents that occurred in Ankara in a year and tried to establish correlated rules via association rule analysis with the help of SPSS Clementine 12.0 program. In the first part of the work, the pre-processing of the data, the proper arrangement of the accident data was provided. In the next part, outlier values were determined and removed and data types were converted to binary type. In the last part, they performed rule extraction using the application of the Apriori algorithm and evaluated the obtained rules.

Çalışkan et al. (2020) used the Apriori algorithm and the FP-Growth algorithm in their study, and performed association rules analysis with the help of the R program on the data set consisting of crime data from July 2016 to April 2018 in the US state of Maryland.

Waterson et al. (2016) used the system-based systematic accident analysis technique developed by Rasmussen, which can be used for accident analysis. With this technique, they examined the change of accidents over time.

Ona et al. (2013) investigated the traffic accidents that occurred on rural highways in the Spanish province of Granada between 2005 and 2008 and tried to determine the severity of damage in traffic accidents by using latent class cluster analysis and Bayesian network techniques together.

Nahar et al. (2013) used three different association rule mining algorithms in their study. They provided association rule extraction on the heart disease data they used. They performed data mining analysis. They found critical risk factors for heart disease for men and women.

Brossette et al. (1998) used infection control data in their study, and examined them in six-month, three-month and one-month time periods. By performing association rules analysis with these data, they produced up to 20,000 association rules.

Association Analysis and Apriori Algorithm

Analyzing large data sets, separating and filtering the patterns throughout the information discovery process and making them ready for the next step is also a part of this process. No data mining algorithm is beneficial, no matter how effective, if the characteristics of the work and data on which the analysis is carried out are not known (Savaş, Topaloğul and Yılmaz, 2012: 7).

Data mining usage areas include database analysis, decision support, identification of similarities between customers, market basket analysis, cross-market sales analysis, competition analysis, customer credit risk research, optimal use of corporate resources, determination of customers' purchasing patterns, finding connections between customer demographics, increasing the response rate in mail campaigns and evaluation of loan requests (Sivri, 2015: 4).

Cross-Industry Standard Process for Data Mining (CRISP-DM)

The cycle in Figure 1 is commonly followed in data mining processes. This cycle is referred to as CRISP-DM. The developers of the process, which is expressed as Turkish cross-industry data mining process, are Daimler-Chrysler (later Daimler Benz AG, Germany), NCR Systems Engineering Copenhagen (Denmark), SPSS (England) and OHra Verzegeringen en Bank Groep BV (Netherlands) (Kücüksille, 2009).

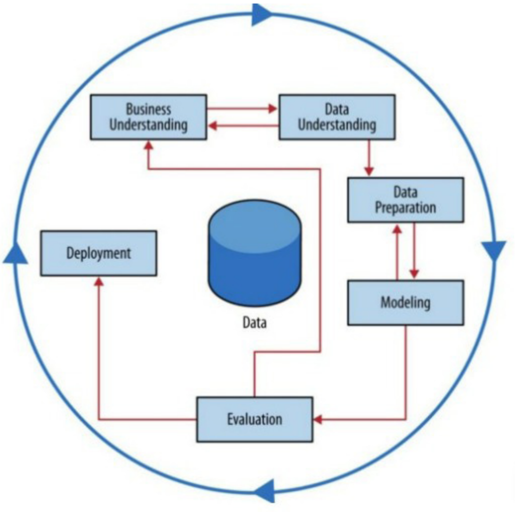

Figure 1. Data Mining Process (Chapman et al., 2000).

The beginning of the data mining process starts with understanding the business. This is the part that needs to be analyzed for the progress of the process. When the process is completed, the goals that are planned to be achieved should be expressed clearly and precisely.

The second stage is understanding the data. A close familiarity with the database will make it possible to process the data analysis to produce the desired results. Missing and incorrect data in the database should be thoroughly analyzed. After this analysis phase, the data preparation phase begins. The data should be converted into a format ready for the algorithms and software to be used, and if there are incomplete or missing data, it should be cleaned.

The modeling stage is the part where the process is technically implemented. In this section, the selection of the modeling technique that is planned to be applied, the creation of the design of the model and the evaluation of the model are carried out.

Bringing the process into report format and documenting it is the final stage. Considering that the process is carried out by the research firm, this part is an important step for the companies that purchase the service.

Properties of Simulation Environment

Association analysis is used as an analysis technique that describes the strength of the relationship between pairs of products purchased together and identifies patterns of co-occurrence. Association analysis is applied in data mining as a method generally used to predict the purchasing behaviors of customers (Sagin, 2018: 10-19).

In association analysis, the aim is to find the relationships between the parts of the set (number set, word set). The relationships found can be used to increase the firm's profits or to make scientific predictions. If it is known that those who buy product X and Y also buy product Z at a high rate, if a customer buys product X and Y but does not buy product Z, that customer is a potential customer Z. Thus, product Z can be advertised to potential customers of product Z. In addition, if the probability of X being together with Y is high and Y is found in a region, then X should be sought in the region (Alpaydin, 2000).

The association analysis technique simply reveals which products are bought together, which products should be included in which campaign, and consumer behavior. It analyzes product data to determine the proximity of product combinations with each other. Based on the result of this analysis, managers can plan and apply in order to implement more effective sales strategies and develop promotional offers to customers (Alan, 2016: 46).

Sales analysis built by association rules is used as a solution to many problems, such as customer purchasing habits, which products will be discounted, how catalogs will be designed, how products on the shelf will be arranged, etc. (Esen, 2009: 33).

Terms Used in Association Rules

The products purchased by the customers, all the transactions performed and the set of objects bought together during these transactions constitute the association analysis. There are some important terms used when applying this analysis, and they are as follows (Dolgun, 2006: 36): antecedent as representing the left side of the rule, antecedent support as the probability of seeing the first product alone, consequent as the result expressing the right side of the rule, consequent support as the support value probability of seeing the second product alone, support as the confidence value probability of seeing two products together, confidence as the probability of selling the second product when the first product is sold, leverage or lift as the value of how much the probability of selling the second product increases when the first product is sold, minimum support as the minimum support value specified, minimum confidence as the minimum confidence value specified.

Association analysis is based on the calculation of various probability measures. The three main metrics covered are support, confidence and lift (Aksoy, 2019: 73). The letters X, Y and N that we will use to define these metrics are explained below.

- “X: first product, Y: second product and N: total purchase”
- **X&Y:** Represents the number of times X and Y are purchased simultaneously. There may be products other than X and Y in the customer’s shopping cart.
- **X⇒Y:** Represents the situation of *those who get X, get Y*. The difference from X&Y is that in special cases, when X and Y are swapped, the result also changes (i.e., Y⇒X). As an example, X and Y are taken as five; when X and Y are analyzed separately, ‘X’ might be purchased 7 times and ‘Y’ is 12 times.
- **Support:** It is the simultaneous occurrence of X and Y in the entire data set. It is the probability of seeing X and Y together (Bhasin, 2020).

  Support = (Simultaneous purchase(X&Y)) / (Total purchase(N))
- **Confidence:** Probability that product Y will be sold when product X is purchased. It is between 0 and 1 (Dolgun, 2006: 37-38).

  Confidence = (Simultaneous purchase (X&Y)) / (Individual purchase (for X))

The high percentage of Confidence and Support values indicates that association rules are reliable. Finding association rules from common object sets provides the smallest support (minimum_support or min_support) and the smallest confidence (minimum_confidence or min_confidence) (Döşlü, 2008: 27-28).

- **Lift (Leverage):** It is applied to find out how much they are purchased together if X and Y are statistically independent. It is used to conclude the rate of increase in purchases of product Y, in between the customers of X. If the lift value is greater than 1, the variables can be dependent on each other and rules can be created about the variables. The result of lift answers the following question; “If there is a relationship, how many times could those who buy product X affect the sales status of product Y?" Lift value is between 0 (zero) and ∞ (infinite) (Han and Kamber, 2006: 266-272).

Lift = Support(X□Y) / (Individual purchase (X)* Individual purchase (Y))

Results obtained with the association analysis may affect the decisions to be made in the planning of the sales process, the catalog and the shelf image, because, while calculating in association analysis, support and confidence criterion are taken as basis in the connection between goods and services. Association rules calculated according to the level of support and confidence indicate the potential relationships of the data. A strong association rule has great support and a high level of confidence (Aras, 2008: 27).

There are two important aspects for association rules to be useful: the rule is relevant to the subject, and the rule is understandable. Understandability and clarity have always been the strength of association rules. The reason for this is that association rules are symbolic and intuitive. In association rules, there are many steps that the number and type of rules can control (Döşlü, 2008: 27).

Association rule is a rule-based machine learning technique for finding relationships in large databases. With this approach, rules are also produced for new analyses, since large amounts of data are analyzed. Apriori algorithm, as a method used to reveal product associations from user shopping, provides the opportunity to see the associations of products purchased according to a threshold value to be determined. Substantially, it has an iterative nature on the basis of the Apriori algorithm (Han and Kamber, 2006: 1-35). It is used to discover frequent item sets in databases containing departure points.

Types of Association Rules

The types of association rules can be classified in many different ways and are listed under three main headings as follows (Döşlü, 2008: 31):

- According to the types of values used in the rule: If the rule is about associations between the presence and absence of objects, this is called a Boolean association rule. These rules are derived from association analysis. If it describes associations between quantitative objects or properties, it is a quantitative association rule. In these rules, quantitative values or properties for objects are divided into ranges.
- According to the dimensions of the data contained in the rule: If the attributes or objects in an association rule represent only one dimension, then the rule is said to be a single dimension association rule

- According to various dimensions of association rules (Dolgun, 2006): Association rules analysis can be an extension of correlation analysis. When the data contains various dimensions of association rules, it is called a multi dimension association rule. At the same time, it can be an extension of the “maxpattern” and “frequent closed itemset” analysis.

With the association rules, the products purchased together by the customers are analyzed. It is used to explain the relationships that occur simultaneously. For example, the customer who buys plane tickets for all family members to go on holiday will be able to rent a car in the holiday region with a probability of over 90% (Şimşek, 2006: 53). More than half of the female customers who buy ready meals also buy care products. Firms use this technique to determine the purchasing targets of their customers.

Apriori Algorithm

The most well-known algorithm in association rules extraction is the Apriori algorithm. This algorithm uses the a priori information of common objects, that is, takes the information from the previous step, therefore, the name “Apriori” is derived from the word “prior.” Unlike other algorithms, there is a difference in the way candidate object sets are produced and the selection of candidate object sets to be counted. The Apriori algorithm focuses on this significant point. Apriori creates candidate object sets by combining common object sets formed in the previous transition. Without dealing with the movements in the data pool, it deletes the smaller subsets formed in the previous transit from the data pool (Döşlü, 2008: 34). The Apriori principle states that if an item set is sparse, all its subsets must be sparse as well. The Apriori algorithm is one of the best algorithms available for joint decision making.

The steps of the algorithm are detailed below (Şekeroğlu, 2010):

- In the first step, threshold values are determined in order to compare support and confidence values. It is expected that the results obtained from the analysis will be greater than or equal to the threshold values.
- In the second step, the number of repetitions for the products to be included in the analysis, that is, the support value, is calculated and the database is scanned for this. After the support values are obtained, they are compared with the threshold support value. If the obtained support value of the product is less than the threshold support value, the relevant rows are excluded from the analysis. Only appropriate records are considered.
- When we come to the third step, the products selected in the second step are grouped in pairs and the support values, that is, the number of repetitions,

are obtained. As in the previous step, these values are compared with the support values, and the lines less than the threshold value are excluded from the analysis.

- In the fourth step, the grouping values are gradually increased by one, and the threshold values are compared in groups of three and four. The process continues as long as the number of repetitions is above or equal to the threshold value. At the last stage, after the product group is determined, association rules are produced by looking at the rule support criterion. Confidence criteria are calculated for each of the rules.

The criteria for applying the Apriori algorithm in probability metrics are as follows (Döşlü, 2008: 34):

- The minimum value is determined for the support and confidence value, and those above that value are included in the rule.
- Support values are checked by listing how many times all products are purchased one by one.
- Binary groups are formed with the remaining products. The same process is applied to the binary groups, and after the ones below the support value are removed, the confidence value of the remaining ones is checked. Those with a confidence value higher than the minimum are chosen as association rules.

Implementing Association Analysis with Data Mining Technique

The data used in the applied study were one-day shopping data in a five-day time period of a company operating in the retail sector in Turkey. The association analysis of the products was carried out with the help of Python by examining 6 categories of products over the dataset, which includes 1,528 customer (single) transactions, including customers who choose one or more products, out of 2,520 invoice transactions.

The variables used in the data are shown in Table 1 below. The main features of the application are as follows:

- Includes shopping information between January 4, 2021 (Monday) and January 8, 2021 (Friday).
- It consists of 6 categories of outerwear, youth wear, casual, women's outerwear, special collection, and sportswear.
- Consists of 1,528 different customer numbers.

Table 1

Variables Used in the Data Set

Variable Name	Definition
Invoice Number	The unique number of each transaction, namely the invoice.
Product Group	Information about category names of products.
Invoice Date	Shows when the product was being purchased.
Price	Shows how much was paid for the product or products purchased.
Customer number	The unique customer number.

In the Apriori algorithm, it is calculated whether the frequency value (number of repetitions) of each product is found in the data studied. The number of repetitions and support values are calculated. A new table is created with products with a support value equal to or above the minimum support (min_support). The larger the minimum support value, the less co-occurrence of products.

The Apriori algorithm was invoked to generate frequent item sets and the minimum support was pulled to 80%. The Apriori algorithm was applied with a minimum support of 80% in the sample dataset. Thus, the probability of purchasing a minimum of 80% of the purchased products in all sales was examined.

Table 2

Minimum support 80% sample frequency table

Support	Items
0.95	Youth Clothing
0.95	Outerwear
0.95	Casual
0.82	Women's Outerwear

As table 2 presents above, according to the minimum support 80%, when the Apriori algorithm is applied, it is seen that sales are made from four categories of products. These product categories are youth wear, outerwear, casual and women's outerwear. It is seen that the larger the minimum support value, the lower the rate of co-occurrence of products.

Table 3

Association Analysis Result Table

Antecedent	*Results*	*The Probability of Seeing the First Product Alone Support Value*	*Possibility of Seeing Second Product Alone Support Value*	*Support*	*Confidence*	*Lift*
Youth Clothing	Outerwear	0.95	0.95	0.95	1.0	1.044
Outerwear	Youth Clothing	0.95	0.95	0.95	1.0	1.044
Casual	Outerwear	0.95	0.95	0.95	1.0	1.044
Outerwear	Casual	0.95	0.95	0.95	1.0	1.044
Women's Outerwear	Outerwear	0.82	0.95	0.82	1.0	1.044

According to Table 3, the results of the association analysis in shopping are given below:

- Products purchased from the youth clothing and outerwear categories are seen in 95% of all purchases.
- Products purchased from women's outerwear and outerwear categories are seen in 82% of all purchases.
- The probability of seeing a product from the youth clothing category alone is 95%.
- The probability of seeing a product taken from the category of women's outerwear alone is 82%.
- In 95 out of 100 purchases, products from the youth clothing and outerwear categories are purchased together.
- In 82 out of 100 purchases, women's outerwear and outerwear products are purchased together.
- The sales of the casual category product increase 1.044 times in outerwear purchases.

Discussions and Conclusions

The way for companies to be protected and survive in the developing competitive environment depends on making use of scientific techniques, especially during the decision phase. One of the most important ways to stand out in a competitive environment is to ensure customer satisfaction. It is to determine the expectations and needs of the customers correctly. Companies can increase customer satisfaction with their efforts to protect their existing customers and gain their customers' loyalty. Increasing customer satisfaction and gaining potential customers is possible with a correct understanding of customer purchasing habits. One of the most important steps in becoming a customer-oriented company is to design the stores in a customer-friendly way. Customer likes and habits can be determined by monitoring their purchasing movements. Based on the data obtained, answering questions such as which product is sold with which product, and whether there is a close relationship between these products being sold together, can be important for companies in terms of developing customer-oriented strategies. In the campaigns to be designed by companies, it is possible to determine which product groups should be directed to which group of customers, in which category, and which product groups should be brought to the fore during seasonal changes. The customer-oriented store layout and the arrangement of the shelves according to their purchasing habits can be determined by the results that can be obtained from the data. For all these, association rules extraction, which is one of the data mining models, can be used.

Behavior patterns that enable companies to make effective decisions were revealed using data mining in the implementation process. In the study, which was carried out using the Apriori algorithm, ideas were produced about which products were more focused by examining the behaviors of the customers. Also, the data includes a specific period. In future studies, extended research can be done for data belonging to different periods of the year or the whole year.

By using association rules in the study, it can be shown that customers' purchasing habits are determined by finding meaningful relationships between products. A large amount of reliable data (without errors and omissions) is a prerequisite in association analysis, because the quality of the extracted rules primarily depends on the quality of the data. In addition to this type of study, which is carried out on the data obtained from the purchasing preferences of the company in the product category, it is also possible to analyze the forecasts of future purchases by examining all product categories or using the data consisting of the product groups that are desired to be more prominent.

Peer-review: Externally peer-reviewed.
Author Contributions: : Conception/Design of study: H.D., Ö.D.B.; Data Acquisition: H.D.; Data Analysis/ Interpretation: H.D., Ö.D.B.; Drafting Manuscript: H.D.; Critical Revision of Manuscript: H.D., Ö.D.B.; Final Approval and Accountability: H.D., Ö.D.B.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

Hakem Değerlendirmesi: Dış bağımsız.
Yazar Katkısı: Çalışma Konsepti/Tasarımı: H.D., Ö.D.B.; Veri Toplama: H.D., Ö.D.B.; Veri Analizi /Yorumlama: H.D., Ö.D.B.; Yazı Taslağı: H.D., Ö.D.B.; İçeriğin Eleştirel İncelemesi: H.D., Ö.D.B.; Son Onay ve Sorumluluk: H.D., Ö.D.B.; H.D., Ö.D.B.
Çıkar Çatışması: Yazar çıkar çatışması bildirmemiştir.
Finansal Destek: Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

References

Aksoy, A. (2019). *Pazarlama ve müşteri ilişkileri yönetiminde veri madenciliği uygulamaları: pazar sepet analizinde apriori algoritmasının uygulanması* (Yüksek lisans tezi), İstanbul Aydın Üniversitesi, Sosyal Bilimleri Enstitüsü.

Alan, B. (2016). Veri madenciliği ve market veri tabanında birliktelik kurallarının belirlenmesi (Yüksek lisans tezi), Recep Tayyip Üniversitesi Sosyal Bilimler Enstitüsü.

Alpaydın E. (2000). Ham veriden altın bilgiye ulaşma yöntemleri. Boğaziçi Üniversitesi, Bilgisayar Mühendisliği Bölümü, 25-42.

Aras, Ü. (2008). Finansal veri madenciliği (Yüksek lisans tezi), Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Bhasin, R. (2020). ApacheSpark kullanarak keşifçi ve pazar sepeti analizi (Çevrimiçi). https://medium.com/analytics-vidhya/shopper-behavior-exploration-and-market-basket-analysis-using-spark-650656d6a0e1

Brossette, Stephen E., Sprague, Alan P., Hardin, J. Mıchael, Waites, Ken B., Jones, Warren T., Moser, Stephen A., (1998). Data mining in hospital infection control and public health surveillance, journal of the American Medical Informatics Association, 5, 373-381.

Chapman, P., Clinton, J., Kerber, R., Khabaz, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). Adım adım veri madenciliği kılavuzu.

Çalışkan, D., Yıldız, K., Doğan, B., ve Aktaş, A. (2020). Crime data analysis with association rule mining. *Int. Per. of Recent Tech. in App. Eng., 2*(2), 42-50. DOI: 10.292228/porta.1,

De Oña, J., López, G., Mujalli, R., and Calvo, F. J. (2013). Analysis of traffic accidents on rural highways using latent class clustering and bayesian networks. *Accident Analysis and Prevention,* 51, 1-10.

Dolgun, M. Ö. (2006). Büyük alışveriş merkezleri için veri madenciliği uygulamaları (Yüksek lisans tezi), Hacettepe Üniversitesi İstatistik Anabilim Dalı, Ankara.

Döşlü, A. (2008). Veri madenciliğinde market sepet analizi ve birliktelik kurallarının belirlenmesi (Yüksek lisans tezi), Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Esen, F. (2009). Veri tabalarından bilgi keşfi: veri madenciliği ve bir sağlık uygulaması (Yüksek lisans tezi), İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Han, J. ve Kamber, M. (2006). Data mining concepts and techniques, Morgan Kauffmann Publishers Inc., 1-35.

Kalıkov, A. (2006). Veri madenciliği ve bir e-ticaret uygulaması (Yüksek lisans tezi), Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Karagöz, N. E. (2007). Market veri tabanında veri madenciliği uygulaması (Yüksek lisans tezi), İstanbul Ticaret Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Küçüksille E. (2009). Veri madenciliği süreci kullanılarak portföy performansının değerlendirilmesi ve İMKB Hisse Senetleri Piyasasında bir uygulama (Yüksek lisans tezi), Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Isparta.

Nahar J., Imam T., Tickle KS., Chen YP (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert systems with applications,* (40), 1086–1093.

Sağın., A.N. ve Ayvaz., B., 2018. Determination of association rules with market basket analysis: an application in the retail sector. *Southeast European journal of soft computing, 7*(1), 10-19.

Savaş, S., Topaloğul, N. ve Yılmaz., M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi,* (7-8).

Shearer, C. (2000). The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing, 5*(4), 13-23.

Sivri, E.Ş. (2015). Veri madenciliği/e-ticaret için ürün tavsiye sistemi geliştirilmesi (Yüksek lisans tezi), İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü.

Şekeroğlu, S., (2010). Hizmet sektöründe bir veri madenciliği uygulaması (Yüksek lisans tezi), İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 135, İstanbul.

Söylemez, İ., Doğan, A. ve Özcan, U. (2016). Trafik kazalarında birliktelik kuralı analizi: Ankara ili örneği. *Ege Akademik Bakış Dergisi*, 16, 11-20. DOI: 10.21121/eab.2016OZEL24423.

Şimşek, U. T. (2006). Veri madenciliği ve müşteri ilişkileri yönetiminde bir uygulama (Doktora tezi), İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Timor, M., Ezerçe, A., Gürsoy., U.T., 2011. Müşteri profili ve alışveriş davranışlarını belirlemede kümeleme ve birliktelik kuralları analizi: perakende sektöründe bir uygulama. *İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadı Enstitü Yönetim Dergisi,* 128-147.

Waterson, P., Jenkins, D. P., Salmon, P. M., & Underwood, P. (2016) Remixing rasmussen: the volution of accimaps within systemic accident analysis applied ergonomics, 59(B), 483-503.5393 sayılı Belediye Kanunu (2005). Resmî Gazete Sayı: 25874.